

Dithered GMD Transform Coding

Ching-Chih Weng, *Student Member, IEEE*, P. P. Vaidyanathan, *Fellow, IEEE*, and Han-I Su, *Student Member, IEEE*

Abstract—The geometric mean decomposition (GMD) transform coder (TC) was recently introduced and was shown to achieve the optimal coding gain without bit loading under the high bit rate assumption. However, the performance of the GMD transform coder is degraded in the low rate case. There are mainly two reasons for this degradation. First, the high bit rate quantizer model becomes invalid. Second, the quantization error is no longer negligible in the prediction process when the bit rate is low. In this letter, we introduce dithered quantization to tackle the first difficulty, and then redesign the precoders and predictors in the GMD transform coders to tackle the second. We propose two dithered GMD transform coders: the GMD subtractive dithered transform coder (GMD-SD) where the decoder has access to the dither information and the GMD nonsubtractive dithered transform coder (GMD-NSD) where the decoder has no knowledge about the dither. Under the uniform bit loading scheme in scalar quantizers, it is shown that the proposed dithered GMD transform coders perform significantly better than the original GMD coder in the low rate case.

Index Terms—Bit allocation, dithered quantization, geometric mean decomposition, linear prediction, transform coding.

I. INTRODUCTION

TRANSFORM coding has played an important role in various fields of signal processing [5], [9]. In the high rate case, given the input statistics, one of the optimal unitary transforms that yields the maximum coding gain is the Karhunen–Loeve transform (KLT). Recently, the prediction-based lower triangular transform (PLT) [7] was proposed. With the aid of the MINLAB [8] structure, PLT was also shown to maximize the coding gain. In [3], the authors applied the generalized triangular decomposition (GTD) theory to transform coding, and showed that both KLT and PLT are special cases of the GTD transform coders. It was also shown that all GTD transform coders achieve the maximum coding gain with optimal bit allocation. Several new transform coders in the GTD transform coder family were proposed and shown to have some good properties. In particular, the GMD transform coder yields the maximum coding gain with uniform bit allocation, i.e., no bit allocation is needed.

Manuscript received November 20, 2009; revised February 11, 2010. First published February 22, 2010; current version published March 26, 2010. This work was supported in part by the Office of Naval Research under Grant N00014-08-1-0709, and by Caltech. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Adrian Munteanu.

C.-C. Weng and P. P. Vaidyanathan are with California Institute of Technology, Pasadena, CA 91109 USA (e-mail: cweng@caltech.edu; ppvnath@systems.caltech.edu).

H.-I. Su is with Stanford University, Stanford, CA 94305 USA (e-mail: hanisu@stanford.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/LSP.2010.2043887

However, many applications require very low rate transform coders [6]. In these scenarios, the GMD transform coder performs poorly (as can be seen from [3, Fig. 12]). This is mainly due to two reasons. First, in the high rate case, the uniform quantizer acts like an independent additive noise source. This approximation is no longer valid in the low rate case, where the quantization noise strongly depends on the quantizer input. Second, in the middle-PLT part of the GMD quantizer, the quantized data is used for prediction. However, the prediction coefficients themselves are obtained from the unquantized data. The effect of this mismatch is no longer negligible in the low rate regime.

In this letter, we propose two transform coding structures: the GMD subtractive dithered (GMD-SD) transform coder and the GMD nonsubtractive dithered (GMD-NSD) transform coder. These two transform coders solve the two difficulties mentioned above in the low rate case. The first difficulty is solved by using dithered quantization [2]. If the decoder has perfect knowledge of the dither signal, the GMD-SD transform coder can be used. In absence of knowledge of the dither signal at the decoder, we propose using the GMD-NSD transform coder. The dither signal is chosen differently in each case in order for the quantization error to be uncorrelated with the quantizer input. The second difficulty is solved by redesigning the prediction coefficients. The predictors are derived from the second order statistics of the quantized data to accommodate the effect of quantization noise. Based on these approaches, we are able to improve the coding performance significantly in the low rate case compared to the original GMD coder proposed in [3].

This letter is structured as follows. In Section II, we introduce the signal model and briefly review the required knowledge. In Section III, we describe the structures of the GMD-SD and the GMD-NSD transform coders. Section IV provides the numerical simulation to compare the performance of different transform coders. The final remarks are summarized in Section V.¹

II. SIGNAL MODEL AND PRELIMINARIES

In this letter, we consider the transform coding system in Fig. 1. The signal \mathbf{x} is an $M \times 1$ real-valued vector coming from a wide-sense-stationary vector process with covariance matrix \mathbf{R}_x . In the encoder, the signal is first passed through a linear transform \mathbf{T} . The transform domain signal \mathbf{y} is then passed through parallel scalar quantizers. The decoder applies the inverse transform matrix \mathbf{T}^{-1} to the quantized signal $\hat{\mathbf{y}}$ and obtains the reconstructed signal $\hat{\mathbf{x}}$. Such a coding system has been discussed in detail in [4], [5]. The scalar quantizers can be modeled as additive noise sources. The classical assumption of the additive noise model is that the noise is independent of the

¹Like the KLT and the PLT coders, the performances of the dithered GMD transform coders are guaranteed under the assumption that the covariance matrix of the source data is known. When source statistics are unknown or varying, modified versions of the current dithered GMD coder need further investigation. Also, the complexity of the proposed coder needs to be reduced for the possibility of practical use.

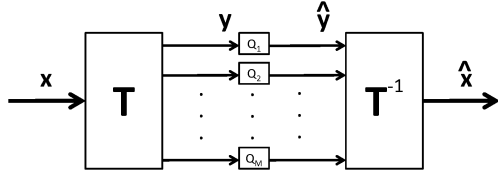


Fig. 1. Schematic of a transform coder with scalar quantizers.

quantizer input, and the variance of the quantization error for a b -bit quantizer can be expressed as $\sigma_q^2 = c2^{-2b}\sigma_y^2$, where σ_y^2 is the quantizer input signal variance, and c is some constant depending on the input signal statistics (see Appendix D.4 of [5]). This assumption approximately holds true when the number of bits b is sufficiently large, and thus it is called the high rate assumption.

Based on the high rate assumption, the optimal transform coder which maximizes the coding gain (defined as the mean square error without coding over the mean square error with coding) is the one that uses linear transform \mathbf{T} which diagonalizes covariance matrix \mathbf{R}_x and some optimal bit loading scheme [5] for the quantizers. This solution is called the Karhunen–Loeve transform (KLT) coder. In [7], the authors propose the PLT coder which exploits linear prediction [10] to design the transform coder. Suppose that the LDU decomposition of \mathbf{R}_x is $\mathbf{R}_x = \mathbf{L}\mathbf{D}\mathbf{L}^T$, where \mathbf{D} is a diagonal matrix, and \mathbf{L} is a lower triangular matrix with unit diagonal entries. The PLT uses $\mathbf{T} = \mathbf{L}^{-1}$ as the linear transform. This amplifies the quantization noise since it is not an orthogonal matrix (see Section IV of [7]). However, with the aid of a variation called the MINLAB structure [8], the PLT coder with suitable optimal bit loading scheme achieves the same optimal performance as the KLT.

Later on, it was observed in [3] that if we apply the PLT to the precoded signal $\mathbf{z} = \mathbf{P}^T\mathbf{x}$ for some orthogonal precoding matrix \mathbf{P}^T , it is possible to achieve the same coding performance as the KLT and the PLT. This observation led the authors in [3] to develop the GTD transform coder family. An example of the GTD transform coder is shown in Fig. 2 for $M = 3$. All instances of the GTD transform coder family are shown to be optimal, and the family includes the KLT and PLT as special cases. Many new optimal transform coders were also developed. The GMD transform coder, which is one instance of the GTD transform coder, in particular achieves the optimal performance without the need of bit loading. Suppose the Cholesky factor of \mathbf{R}_x is \mathbf{A} (i.e., $\mathbf{R}_x = \mathbf{A}\mathbf{A}^T$), the GMD coder first computes the GMD of \mathbf{A}^T , i.e., $\mathbf{A}^T = \mathbf{Q}\mathbf{R}\mathbf{P}^T$, where \mathbf{R} is an upper triangular matrix with identical diagonal entries, then the orthogonal matrix \mathbf{P}^T is used as the precoder.

The nice properties of the GMD coders, however, are no longer observed in simulation results when the bit rate is low (see Fig. 12 in [3]). This is mainly due to the two reasons stated in Section I. The main objective of this letter is to show how to improve the GMD transform coder in the low rate case.

The dithered quantizer is shown to render the quantization noise statistics. Suppose that Q is a scalar quantizer with step size Δ . In a dithered quantizer, the sum of the input signal x and the dither w is passed through the scalar quantizer Q , where the dither is independent of the input. The output signals of subtractive and nonsubtractive dithered quantizers are

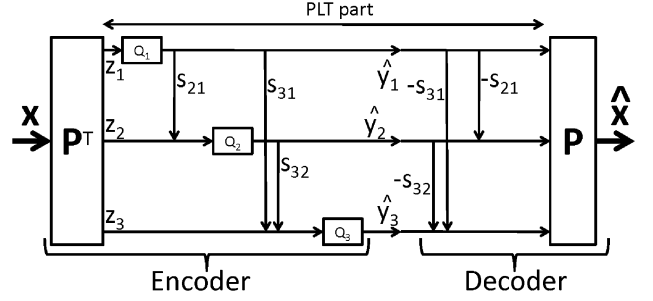


Fig. 2. GTD Transform coder implemented using MINLAB structure.

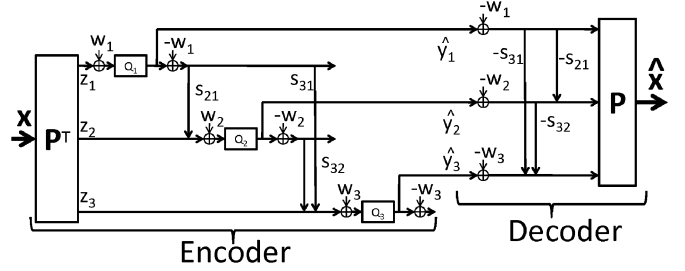


Fig. 3. Subtractive dithered GMD transform coder.

$\hat{x} = Q(x + w) - w$ and $\hat{x} = Q(x + w)$, respectively. Assume that the maximum quantization error is less than $\Delta/2$. It is shown in [1], [2] that (i) if the dither w is uniformly distributed on $(-\Delta/2, \Delta/2]$ then the quantization error of the subtractive dithered quantizer $Q(x + w) - (x + w)$ is independent of the input signal and is uniformly distributed on $(-\Delta/2, \Delta/2]$, and (ii) if the dither w is the sum of $n \geq 2$ independent random variables uniformly distributed on $(-\Delta/2, \Delta/2]$, then the k -th moment of the quantization error of the nonsubtractive dithered quantizer $Q(x + w) - x$ is independent of the input distribution for $k = 1, 2, \dots, n$, and its variance is equal to $(n + 1)\Delta^2/12$. In this letter, we propose to use dithered quantization along with the GMD coder. Also, a design method for the predictors in the PLT structure used in the GMD coder is proposed to incorporate the low rate mismatch.

III. DITHERED GMD QUANTIZER

We consider transform coder structures with dithered quantizers combined with the GMD coder. Fig. 3 shows an example of the structure of a GMD subtractive dithered (GMD-SD) coder with $M = 3$. The dither w_i is added to the input of uniform quantizer Q_i for $i = 1, 2, \dots, M$. After the uniform quantizer, the dither is subtracted from the quantized signal. The resulting signal is then multiplied with the predictor coefficients $s_{1i}, s_{2i}, \dots, s_{Mi}$ for the use of following substream quantizers. The quantized signal \hat{y}_i is then stored or transmitted to the decoder side. The subtractive dither quantizer assumes the decoder has knowledge about the dither signal. In the decoder, the dither is first subtracted. The resulting signal then undergoes the inverse operation of the prediction process and the $M \times M$ matrix \mathbf{P} , which yields the reconstructed signal \hat{x} .

In the nonsubtractive dithered quantizer, the dither knowledge is lacking in the decoder. Fig. 4 shows the structure of GMD nonsubtractive dithered (GMD-NSD) coder. In the encoder of GMD-NSD, the quantized signal is directly multiplied with the

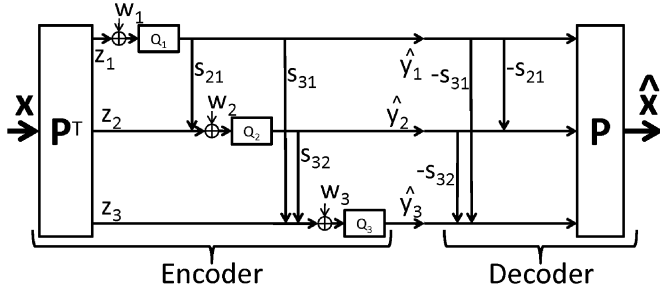


Fig. 4. Nonsubtractive dithered GMD transform coder.

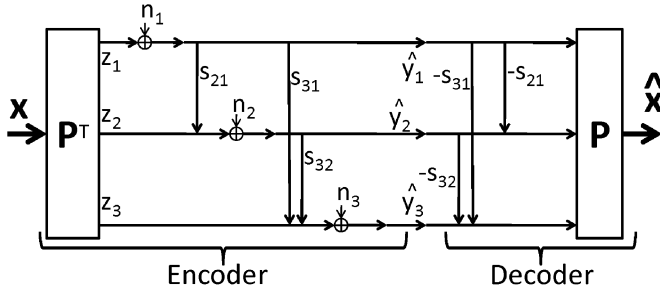


Fig. 5. Equivalent model of dithered GMD transform coder.

predictor coefficients for the use of following substream quantizers, without being first subtracted from the dither. For both cases, the transform coder can be modeled as in Fig. 5, where the i th dithered quantizer is modeled as an additive noise source n_i . Note that the statistics of noise source n_i are different in these two cases, as elaborated at the end of Section II.

In the following, we will use the model in Fig. 5 to design the predictor coefficients. The transformed signal \mathbf{z} is passed through a prediction-based lower triangular transform coder [7] implemented using the MINLAB structure [8]. The resulting encoded signal is $\hat{\mathbf{y}} = \mathbf{S}(\mathbf{P}^T \mathbf{x} + \mathbf{n})$, where $\hat{\mathbf{y}} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_M]^T$, $\mathbf{n} = [n_1, n_2, \dots, n_M]^T$, and the prediction matrix \mathbf{S} is a unit-diagonal lower triangular matrix that consists of prediction coefficients in the MINLAB structure. The covariance matrix of $\hat{\mathbf{y}}$ can be written as

$$\hat{\mathbf{R}}_y = \mathbf{S} \mathbf{E}[(\mathbf{z} + \mathbf{n})(\mathbf{z} + \mathbf{n})^T] \mathbf{S}^T = \mathbf{S}(\mathbf{R}_z + \mathbf{R}_n) \mathbf{S}^T$$

where we assume the noise samples are uncorrelated with the signal \mathbf{z} . Therefore, the MMSE prediction matrix can be obtained by viewing the signal covariance matrix of the middle-part (PLT part) as $\mathbf{R}_z + \mathbf{R}_n$ instead of \mathbf{R}_z . By similar derivation as in Section III of [7], the MMSE lower triangular prediction matrix can be written as

$$\mathbf{S} = \mathbf{L}_1^{-1} \quad (1)$$

where \mathbf{L}_1 is the unit-diagonal lower triangular matrix in the LDU decomposition of $\mathbf{R}_z + \mathbf{R}_n$:

$$\mathbf{R}_z + \mathbf{R}_n = \mathbf{L}_1 \mathbf{D} \mathbf{L}_1^T. \quad (2)$$

Suppose that the noise variance depends only on the quantizer step size, and the noise samples are uncorrelated with each

other.² We will first assume that the step size of each quantizer can be made the same (same step-size rule). This implies the signal substream to each quantizer has the same variance, or equivalently $\hat{\mathbf{R}}_y = d\mathbf{I}$, where d is some constant. Later we will prove that this is possible without bit allocation, but with the aid of properly designed precoder \mathbf{P}^T . Under this same step size assumption, the noise covariance matrix $\mathbf{R}_n = \sigma^2 \mathbf{I}$, where σ^2 is the noise variance that depends on the step size. With the use of prediction matrix \mathbf{S} in (1), the covariance matrix of the encoded signal $\hat{\mathbf{y}}$ is $\hat{\mathbf{R}}_y = \mathbf{D}$, where \mathbf{D} is a diagonal matrix. The question now is whether there exists an orthogonal matrix \mathbf{P} so that $\hat{\mathbf{R}}_y = d\mathbf{I}$. The following theorem asserts the existence of such \mathbf{P} .

Theorem 1: Suppose \mathbf{A} is some Cholesky factor of \mathbf{R}_x , i.e., $\mathbf{R}_x = \mathbf{A} \mathbf{A}^T$. Consider the geometric mean decomposition

$$\begin{bmatrix} \mathbf{A}^T \\ \sigma \mathbf{I} \end{bmatrix} = \mathbf{Q} \mathbf{R} \mathbf{P}^T \quad (3)$$

where \mathbf{R} has equal diagonal entries $\mathbf{r} = [r \ r \ \dots \ r]$. If the precoder \mathbf{P}^T in Fig. 5 is taken as the one in (3), then $\hat{\mathbf{R}}_y = \mathbf{D} = d\mathbf{I}$ for some constant d .

Proof: From (2), by using the MMSE prediction matrix $\mathbf{S} = \mathbf{L}_1^{-1}$, the covariance matrix $\hat{\mathbf{R}}_y = \mathbf{D}$. Thus we only need to prove that the LDU decomposition of $\mathbf{R}_z + \sigma^2 \mathbf{I} = \mathbf{P}^T \mathbf{R}_x \mathbf{P} + \sigma^2 \mathbf{I}$ has $\mathbf{D} = d\mathbf{I}$. To prove this, observe that

$$\begin{aligned} \mathbf{P}^T \mathbf{R}_x \mathbf{P} + \sigma^2 \mathbf{I} &= \mathbf{P}^T \begin{bmatrix} \mathbf{A} & \sigma \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}^T \\ \sigma \mathbf{I} \end{bmatrix} \mathbf{P} \\ &= \mathbf{P}^T \mathbf{P} \mathbf{R}^T \mathbf{Q}^T \mathbf{Q} \mathbf{R} \mathbf{P} \\ &= \mathbf{R}^T \mathbf{R} = \mathbf{L}_1 \mathbf{D} \mathbf{L}_1^T \end{aligned}$$

where \mathbf{L}_1 is taken as $\mathbf{r}^{-1} \mathbf{R}^T$, and $\mathbf{D} = \mathbf{r}^2 \mathbf{I}$. This completes the proof. ■

This theorem suggests a method for finding the precoder \mathbf{P}^T . With such \mathbf{P}^T as the precoder and \mathbf{S} in (1) as the prediction matrix, we are able to have $\hat{\mathbf{R}}_y = \mathbf{r}^2 \mathbf{I}$. Therefore, the scalar quantizers of the dithered GMD transform coder can use uniform step size without bit allocation.

The design procedure for the GMD-SD (or GMD-NSD) transform coder, given the input covariance matrix, is summarized in the following.

- 1) Determine the uniform quantizer step size Δ according to the bit rate.
- 2) Determine σ from the quantizer step size, e.g., $\sigma = \sqrt{\Delta^2/12}$ for GMD-SD using the uniform pdf dither, and $\sigma = \sqrt{\Delta^2/4}$ for GMD-NSD using the triangular pdf dither.
- 3) Compute the Cholesky factor \mathbf{A} of \mathbf{R}_x : $\mathbf{A} \mathbf{A}^T = \mathbf{R}_x$.
- 4) Compute the geometric mean decomposition as in (3).
- 5) Compute the LDU decomposition: $\mathbf{P}^T \mathbf{R}_x \mathbf{P} + \sigma^2 \mathbf{I} = \mathbf{L}_1 \mathbf{D} \mathbf{L}_1^T$, and take $\mathbf{S} = \mathbf{L}_1^{-1}$.
- 6) Construct the transform coder structure using \mathbf{P}^T and \mathbf{S} as in Fig. 3 and Fig. 4 for GMD-SD and GMD-NSD transform coder, respectively.

The complexity of the successive decompositions in the above algorithm is in the same order as that in the GMD transform coder described in [3]. A complete comparison of the

²This is achievable by using a uniformly distributed dither for subtractive dithered quantizer and a triangular-pdf dither for nonsubtractive dithered quantizer (see Section II and [1], [2]).

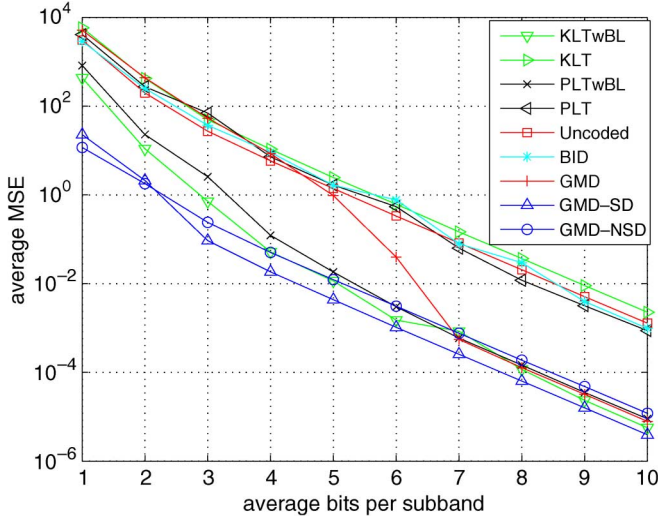


Fig. 6. Performance of different transform coders.

design and implementation costs of different transform coders can be seen in Table I in [3].

IV. NUMERICAL EXAMPLE

In this section we provide numerical comparisons of several transform coders. The signal \mathbf{x} is generated by a zero mean Gaussian vector process with prescribed covariance matrix \mathbf{R}_x . The number of data streams $M = 8$ in the experiments. Uniform roundoff quantizers are assumed. Each quantizer adapts its step size according to the variance of the Gaussian input ([9, p. 818]). For each case, we run the Monte Carlo simulations for calculating the arithmetic mean of mean square error (AM-MSE). In each trial, we first generate the input covariance matrix by multiplying a fixed diagonal matrix $\mathbf{\Lambda}_x$ with a randomly generated orthogonal matrix on the left and its transpose on the right. The input vector \mathbf{x} is then generated according to this covariance matrix. Fig. 6 shows the AM-MSE performance of different transform coders for $\mathbf{\Lambda}_x = 10^{-5} \times \text{diag}([10^7 \ 10^6 \ \dots \ 10^0])$. The following five methods use the same number of bits in each of their quantizers: “KLT” which uses Karhunen–Loeve transform; “PLT” which uses the original structure proposed in [7]; “Uncoded” which directly quantizes the signal; “BID” which uses the bidiagonal decomposition; “GMD” which uses the geometric mean decomposition. The KLT and the PLT with optimal bit allocation [3] but no dithering (represented as “KLTwBL” and “PLTwBL,” respectively) are also simulated for comparison. It should be noted that it is claimed in [3] that under the high bit rate assumption, “GMD,” “KLTwBL,” and “PLTwBL” coders achieve the same minimum AM-MSE performance.

The performances of the two new structures, “GMD-SD” and “GMD-NSD,” are also shown in Fig. 6. In “GMD-SD,” the dither signal is generated from the uniform pdf that satisfies Schuchman’s condition [1]. In particular, it exhibits a variance of $\Delta^2/12$, where Δ is the step size. In “GMD-NSD,” the dither signal is generated from triangular-pdf (Section III-C. in [2]). This dither pdf renders both the first and second moments of the total error independent of the quantizer input. In particular, it is the unique choice of zero-mean dither pdf which renders

the first two moments of the total error independent of the input while minimizing the second [2]. The variance of the error is $\Delta^2/4$. These two dithered GMD quantizers are designed by the method described in Section III.

From Fig. 6, we see that in the low rate regime, the two proposed dithered transform coders have better performance than all the other transform coders. The two optimal-bit-allocated coders, “KLTwBL” and “PLTwBL,” also perform worse than the two dithered transform coders due to the collapse of the high rate assumption. It can also be seen that in the extremely low rate case (one and two bits), there is a performance degradation in “GMD-SD” and “GMD-NSD”. This is because at such low rate, the step size of the dither signal is too large, which makes the chance of overflow much higher for Gaussian sources and violates the nonoverflow assumption in the dither quantizer theory [1]. In the high rate regime, the two proposed dithered GMD transform coders perform comparably with the three coders (“KLTwBL,” “PLTwBL,” and “GMD”) which are designed under the high rate assumption. The AM-MSE of “GMD-NSD” is approximately three times of the AM-MSE of “GMD-SD.” For fixed AM-MSE, “GMD-NSD” needs to have about one more bit than “GMD-SD” needs. The AM-MSE of “GMD” is between those two. The results of this example suggest the following: if the dither signal is available both at the decoder, then “GMD-SD” is a better candidate since it has the lowest AM-MSE in the low rate and high rate regimes; if the dither is not available in the decoder, then we should use the “GMD” coder in the high rate regime, and use “GMD-NSD” in the low rate regime.

V. CONCLUSIONS

In this letter, we described two new transform coder structures: the GMD-SD and the GMD-NSD coders. Both coders are designed without bit allocation, and use the same number of bits and the same step size in their uniform quantizers. It is shown that the proposed coders perform significantly better than the existing GMD transform coder in the low rate case.

REFERENCES

- [1] R. M. Gray and T. G. Stockham, “Dithered quantizers,” *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 805–812, May 1993.
- [2] R. A. Wannamaker, S. P. Lipshitz, J. Vanderkooy, and J. N. Wright, “A theory of nonsubtractive dither,” *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 499–516, Feb. 2000.
- [3] C. C. Weng, C. Y. Chen, and P. P. Vaidyanathan, “Generalized triangular decomposition in transform coding,” *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 566–574, Feb. 2010.
- [4] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer, 1992.
- [5] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
- [6] S. Mallat and F. Falzon, “Analysis of low bit rate image transform coding,” *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 1027–1042, Apr. 1998.
- [7] S. M. Phoong and Y. P. Lin, “Prediction-based lower triangular transform,” *IEEE Trans. Signal Process.*, vol. 48, no. 7, pp. 1947–1955, Jul. 2000.
- [8] S. M. Phoong and Y. P. Lin, “MINLAB: Minimum noise structure for ladder-based biorthogonal filter banks,” *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 465–476, Feb. 2000.
- [9] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [10] P. P. Vaidyanathan, *The Theory of Linear Prediction*. San Rafael, CA: Morgan & Claypool, 2008.